

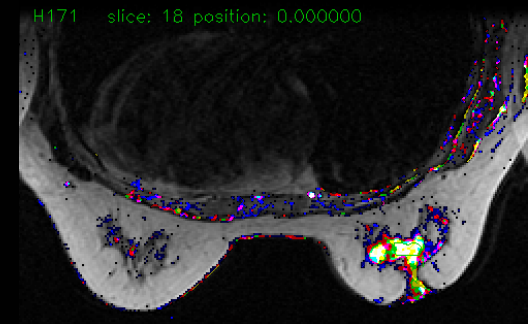
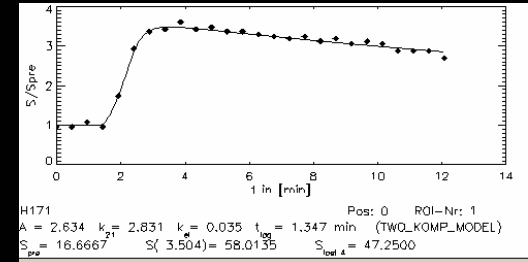
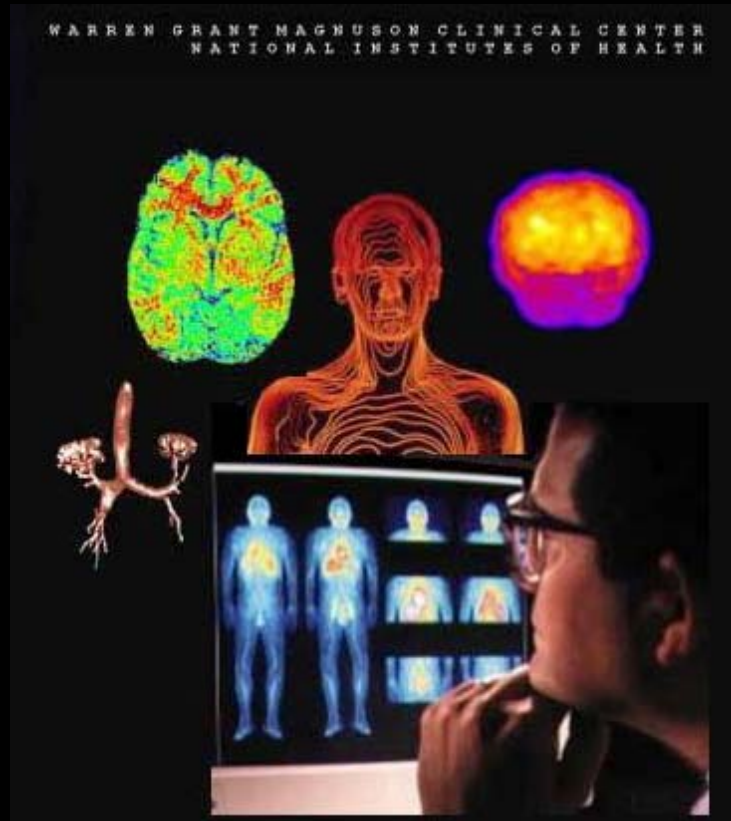
Cancer Imaging Informatics: Synthesis, Summary and Discussion

Ronald M. Summers, M.D., Ph.D.

Clinical Center
Diagnostic Radiology Department
National Institutes of Health
Bethesda, MD

www.cc.nih.gov/drd/summers.html
Biomedical Imaging Program

NIH Clinical Image Processing Service



Overview

- Review of the Meeting
- Perspective of a Clinical Researcher
- Challenges for the Future

Acronym Overload!



DMIST IND
RCET MIRC CIT
BIRN IHE
XML ACRIN NDA
QARC WEAR LIDC
LONI XSL
SQL OMIM
NDMA RTOG

Common Themes

- Standards
 - DICOM (VL, SR)
 - XML
 - HL-7
 - CDE (Common Data Elements)

Common Themes

- Security and confidentiality
 - Smart cards
 - Https
 - Audit trail; “Who saw my records and when?”
 - HIPAA
 - Informed consent to make patient data public (should be requested in ALL consents)
 - Encryption
 - Anonymizing data

Common Themes

- Web based
 - Terraserver
 - SkyServer
 - NCBI databases
(GENBANK, OMIM, ...)
 - CIT/NINDS Stroke Project
 - Dermatology Atlas
 - LIDC

Common Themes

- Big Data
 - Digital Mammographic Imaging Screening Trial
 - National Digital Mammography Archive
 - Lung Image Database Consortium
 - Radiotherapy Databases
 - Dermatology Atlas
 - Cancer Therapy Evaluation Program
 - BIRN

Common Themes

- Faith in the principle that “new tools lead to new discoveries”
- Faith in technological solutions
- Software tools

Common Themes

- Connectivity is good (Value increases $>$ linearly)
- Bigger is better
- Annotation and metadata fundamental

NIH NCBI



National Center for Biotechnology Information

National Library of Medicine

National NCBI banner health

PubMed

Entrez

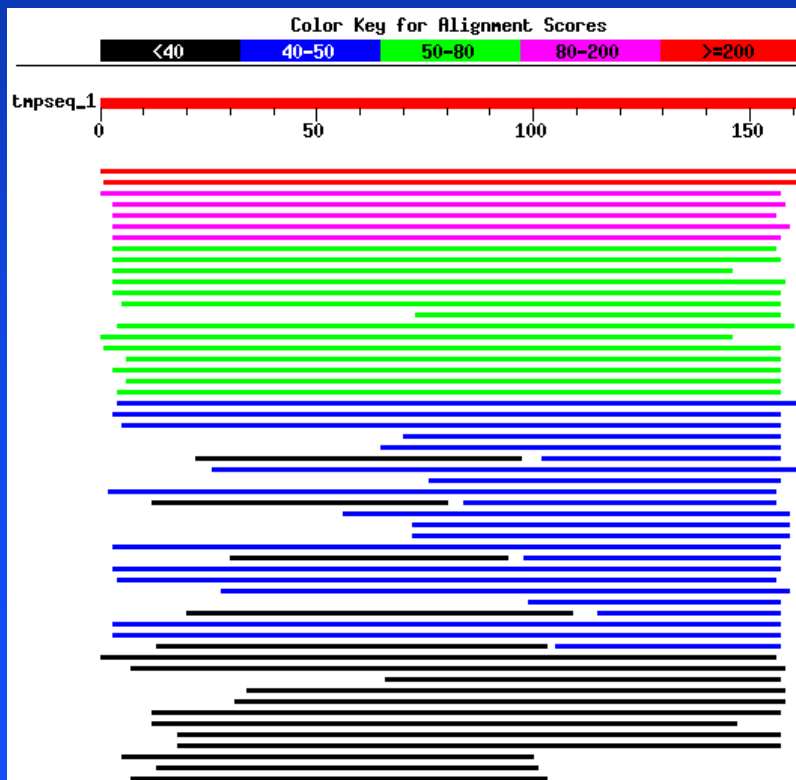
BLAST

OMIM

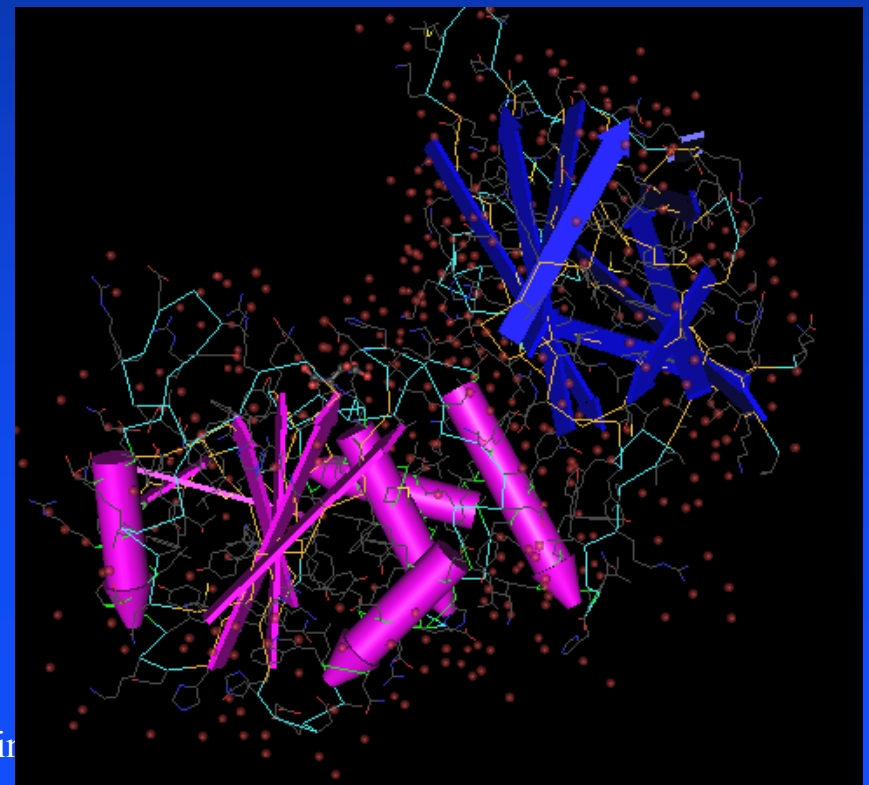
Books

TaxBrowser

Structure



Medical Imagin



Need To Know More

- Emphasis on infrastructure
- Many of the projects we heard about are “big data” (“big science”); expensive
- Results? Proof that patient care or basic research has improved?
- What insights were gleaned? (Science and clinical problems drive design of the informatics, not vice versa)
- Are the results generally applicable?

Need To Know More

- Ease of use
- Training
- Changing closely held patterns of behavior
(The PI who doesn't want to let go of the data)

Impressions

- Huge amount of biomedical information freely available on the web (journal articles, sequences, radiologic images, photographs, histology, etc.)
- Software and tools are an entirely different story

Impressions

- Impressive software was shown
- Is the software mature enough to trust it? How do we know (validation)?
- “Proof of concept”
- Why isn’t more of the software freely available? (exceptions: DICOM, MIRC)
- Duplication of effort
- Open software movement

Please...

- Don't forget the small research projects
- NIH R01's support many research groups consisting of a PI, post-docs and grad students
- As fundamental to research progress as small businesses are to sustaining our economy
- Small research projects drive much scientific innovation, are often more strapped for resources, yet still may generate large datasets

Patterns of Image Processing Research

- Phantom study
- Small clinical trial (proof of concept)
- Medium size clinical trial (first level validation of concept in new dataset)
- Large clinical trial (statistical significance, stratification of patient population)

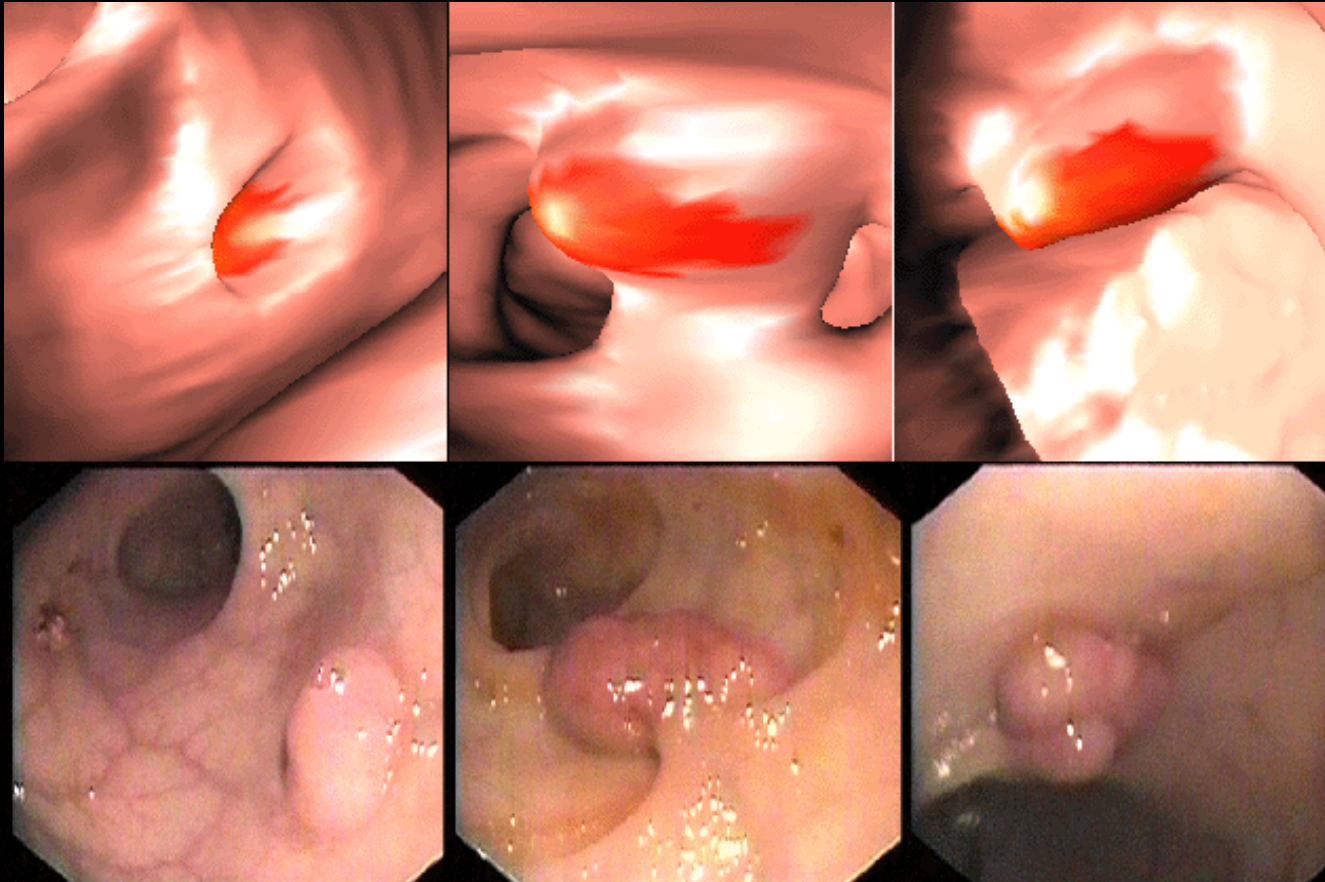
Patterns of Image Processing Research

- Technical development ongoing throughout
- Examples: CT Colonography

CT Colonography

- Minimally invasive alternative to conventional colonoscopy
- Detects colonic polyps, the precursor lesion to colon cancer
- CT scanning
- Bowel prep

CT Colonography CAD

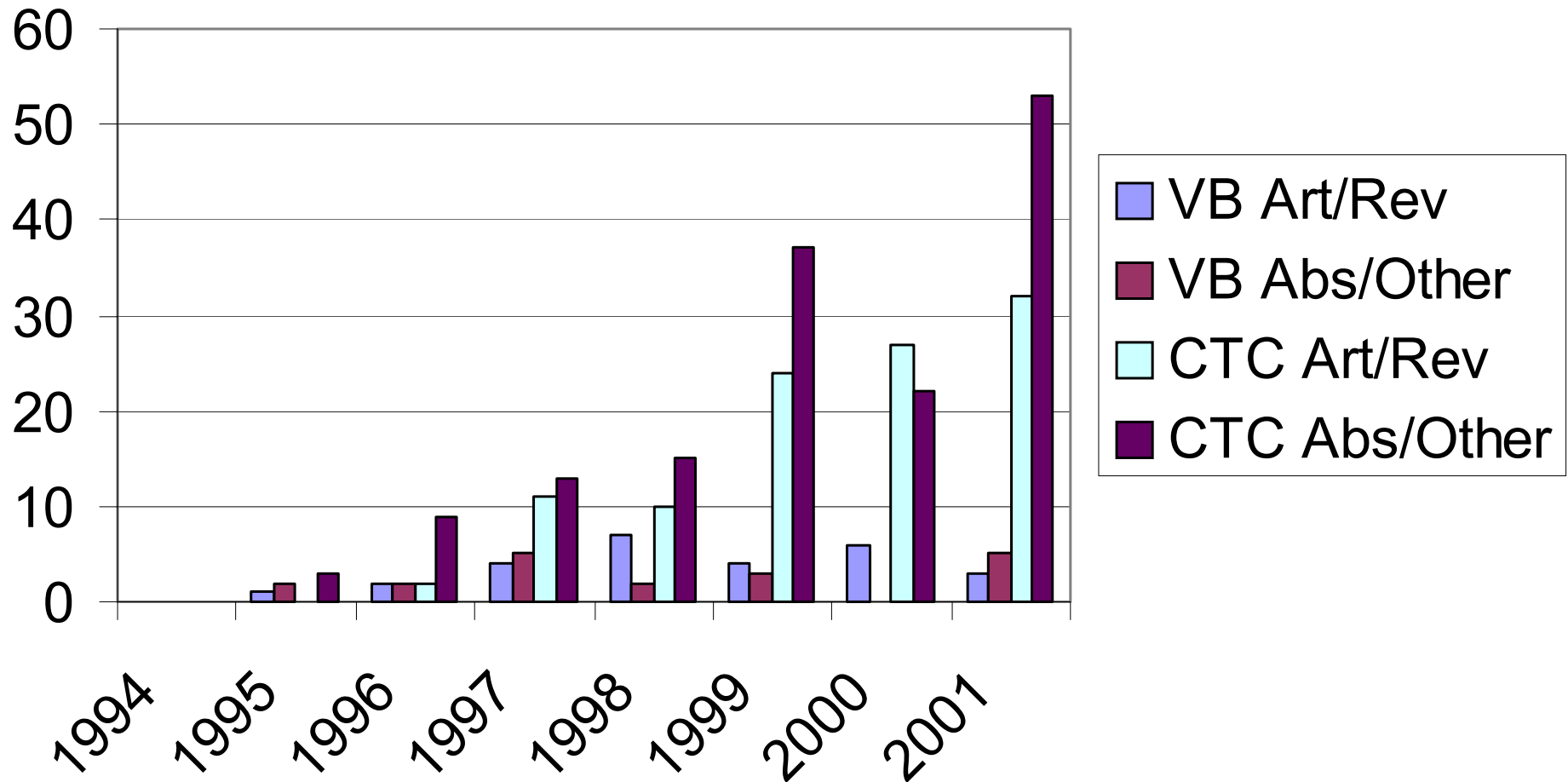


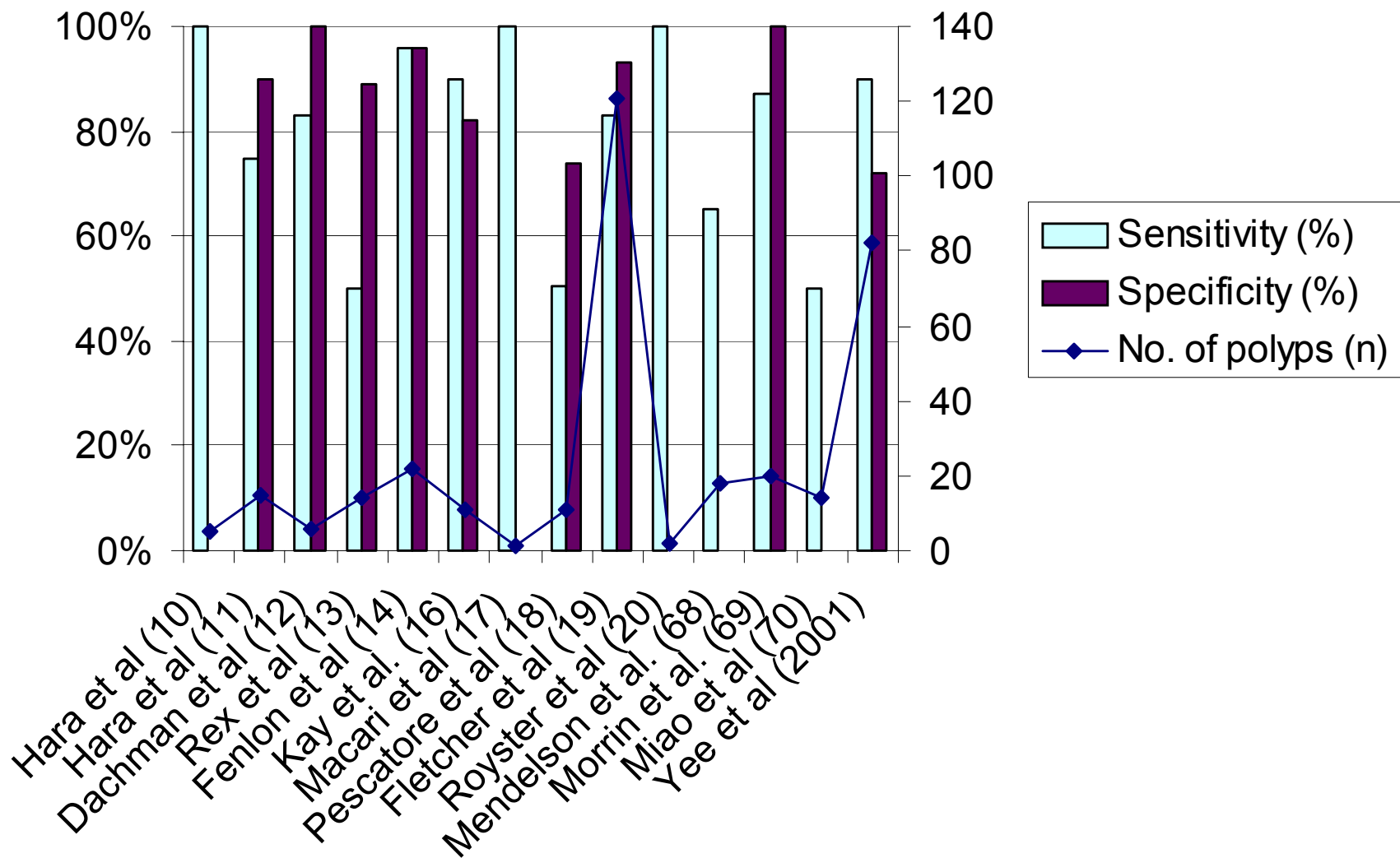
Three polyps in the sigmoid colon of a 68
year old male (1.0, 1.5, 1.0 cm)

R. Summers, D. Johnson et al. Radiology 2001



Research Progress in CT Colonography





Current Status of CTC CAD

- Research in preliminary stage of development
- Early clinical trials at several academic centers
- Further development hindered by lack of large well-validated datasets

CTC Image Database

- Largest CAD studies had 20 polyps
- 1 cm polyps suitable for CAD in 8% or fewer of subjects
- Takes about 1 year to get 20 patients with 1 cm polyps to enroll in study

CTC Image Database

- 200 studies from Mayo Clinic
- 1500 studies from NNMC consortium
- MS Access database containing supporting information (colonoscopy report, coordinates of polyps, pathology report, clinical data)
- Extremely time consuming and labor intensive
- Data handling and annotation tools beneficial

CTC Image Database

- Multiple institutions with CAD expertise are accruing their own databases, some with difficulty

ACRIN CTC Image Database

- 200 CTC studies of patients having polyps 1 cm or larger
- Well annotated – by 3 experienced reviewers
- From multiple institutions
- Side benefit of a larger \$1M clinical trial
- Distribution method still evolving – by application?

ACRIN CTC Image Database

- Final dataset will be very large
- 50 GB compressed
- Sneaker net delivery

Suggestions

- Based upon my experience, I have some suggestions...

Benefiting the Public Good

- Lack of access to data is a common problem
- Studies with large “N” have better statistics, generally have greater clinical impact
- Developing large clinical databases is expensive, time consuming, difficult
- Case material may be limited or localized
- Brain power to analyze the data or come up with new ideas may be more geographically diverse

Data Sharing Culture

- Need data and software sharing culture akin to that of the genome research community
- Currently lots of resistance: It's MY data!
- In reality, once you've published it, it's often of little value to you since you don't have the resources to do more with it or the technology used to generate it has moved on

Data Sharing Culture

- Journal editors cannot enforce data sharing unless mechanism in place

How to Implement

- Grant funders: require sharing or points will be deducted (date and means of data/software release, method of quality assurance); enforce through review at renewal

How to Implement

- Convert more grants to contracts – require a deliverable
- Add a “contract component” to a grant to preserve investigator flexibility/independence
- Researchers keep copyright, patent rights

How to Implement

- Develop common ROI/VOI annotation and database format
- Ground truth is fundamental to utility of image databases, must be high quality, explicitly defined and described
- Level of peer review of the data must be explicitly defined

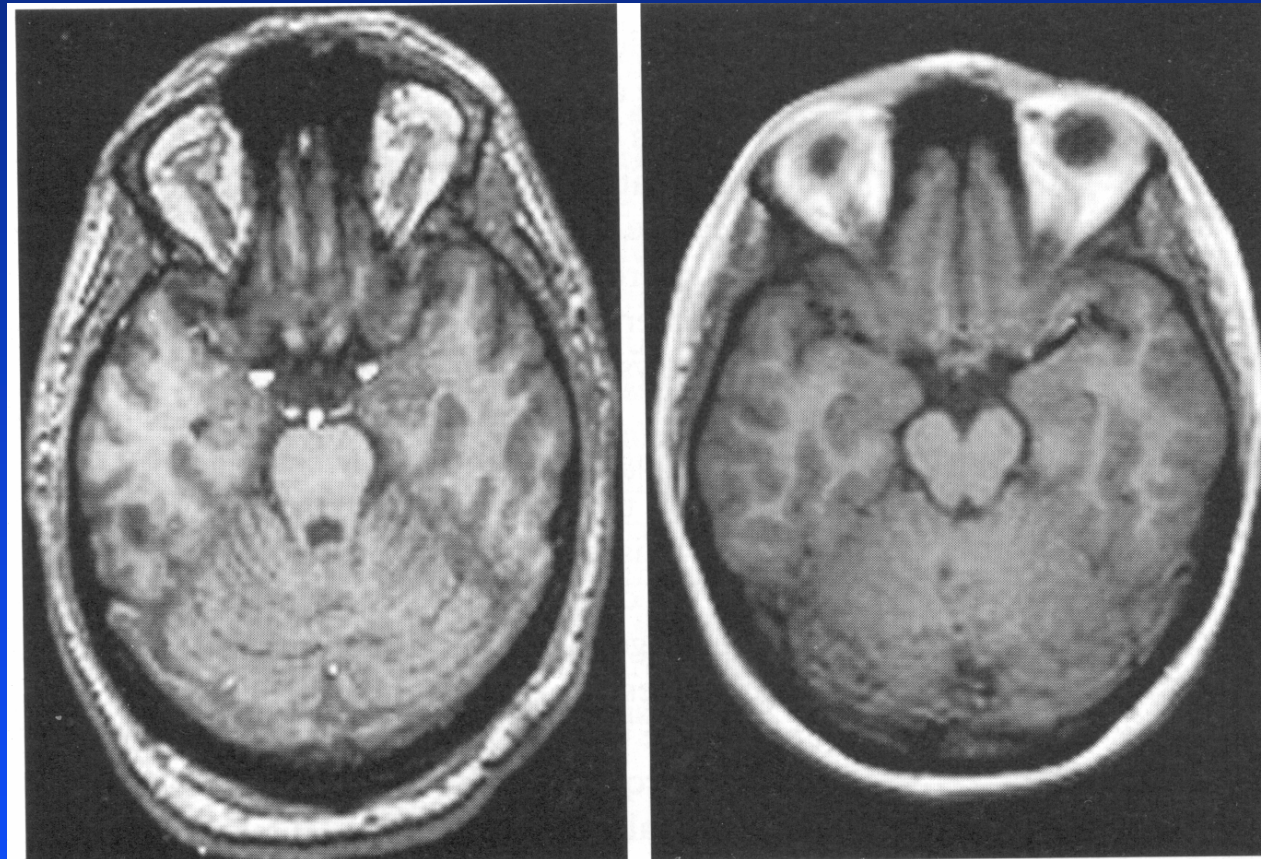
Policy of the Journal **Science**

- “Before publication, large data sets, including protein or DNA sequences and crystallographic coordinates, must be deposited in an approved database and an accession number provided for inclusion in the published paper. Coordinates must be released at the time of publication. Approved databases include ... Other repositories that allow free access to the data for purposes of verification and replication may be acceptable with the approval of the Editor-in-Chief. “

How to Implement

- Radiology and biomedical engineering journal editors should require release of data and supporting software
- NIH could help scientific and clinical community by providing easy and subsidized public repository

Content Based Image Retrieval



Software Repository

- Images alone may not be enough
- Images can be complex
- Software tools may be needed
- Software tools difficult to create quickly from a technical article
- Often requires expertise outside a researcher's specialty

Software Repository

- Scientific goal to reproduce results of others
- Need access to BOTH the data and the tools
- Source code and executable
- Web form to enter requirements, instructions on how to build the source, link to other libraries

Software Repository

- Example: MRI simulation software (MRM 1986)
- Distributed on reel of magnetic tape for \$100
- Could be distributed on the web now for free
- Example: MRA vascular segmentation software (www.cc.nih.gov/drd/software.html); Virtual endoscopy navigation software (www.cc.nih.gov/drd/endoscopy.htm)

Future Challenges

- Change culture and provide infrastructure to make data sharing routine
- Make infrastructure simple for broad spectrum of users (researchers, clinicians)

“Killer App”

- ENDNOTE !?

Future Challenges

- Create software repository to advance state of the art
- Integrate imaging databases with other biomedical databases and into clinical practice

Conclusions

- Projects presented in this workshop were impressive
- Incredible technical skill and foresight
- With encouragement (\$\$\$) and coordination, acceleration of increase of knowledge and improvement of public health